McAllister-Williams RH, Anderson IM, Finkelmeyer A, Gallagher P, Grunze HCR, Haddad PM, Hughes T, Lloyd AJ, Mamasoula C, McColl E, Pearce S, Siddiqi N, Sinha BNP, Steen N, Wainwright J, Winter FH, Ferrier IN, Watson S, ADD Study Team.

Antidepressant augmentation with metyrapone for treatment-resistant depression (the ADD study): a double-blind, randomised, placebo-controlled trial.

***The Lancet Psychiatry* 2016, 3(2), 117-127.**

Antidepressant augmentation with metyrapone for treatment resistant depression (The ADD Study): a double-blind, randomised, placebo controlled trial

R. Hamish McAllister-Williams*[1,2] Ian M. Anderson[4], Andreas Finkelmeyer[1,2], Peter Gallagher[1], Heinz C.R. Grunze[1,2,3], Peter M. Haddad[4,5], Tom Hughes[6], Adrian J. Lloyd[1,2], Chrysovalanto Mamasoula[7], Elaine McColl[7], Simon Pearce[8], Najma Siddiqi[9,10], Baxi N.P. Sinha[11], Nick Steen[7], June Wainright[1], Fiona H Winter[1], I. Nicol Ferrier[1,2], Stuart Watson[1,2], and the ADD Study Team[**]

*** Corresponding Author:**

Academic Psychiatry, Wolfson Research Centre, Campus for Ageing and Vitality, Newcastle upon Tyne, NE4 5PL. Email: r.h.mcallister-williams@ncl.ac.uk

** **The ADD Study Team:** See supplementary material in web appendix.

Affiliations:

1 - Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
2 – Northumberland Tyne and Wear NHS Foundation Trust, Newcastle upon Tyne, UK
3 - Department of Psychiatry and Psychotherapy, Paracelsus Medical University Salzburg, Austria
4 - Neuroscience and Psychiatry Unit, Manchester University, Manchester, UK
5 - Greater Manchester West Mental Health NHS Foundation Trust
6 - Leeds and York Partnership NHS Foundation Trust, Leeds, UK
7 - Newcastle Clinical Trials Unit, Newcastle University, Newcastle upon Tyne, UK
8 – Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK
9 – Bradford District Care NHS Foundation Trust, UK
10 - Institute of Health Sciences, University of Leeds, Leeds, UK
11 – Tees, Esk and Wear Valleys NHS Foundation Trust, UK

Abstract

Background

A significant number of patients with major depressive disorder (MDD) have treatment resistant depression (TRD), defined as failure to respond adequately to two consecutive courses of antidepressants. A Cochrane review suggested that anti-glucocorticoid augmentation of antidepressants may be effective in TRD. The Antiglucocorticoid augmentation of anti-Depressants in Depression (ADD Study) was a multicentre two-arm double-blind, patient randomised, placebo controlled trial of the cortisol synthesis inhibitor, metyrapone, for augmentation of serotonergic antidepressants in patients with TRD in a UK setting.

Methods

A randomised double-blind parallel-group trial of metyrapone 500mg twice daily versus placebo for three weeks, added to on-going antidepressants, in patients aged 18-65 with TRD. On-line computer-generated random allocation was stratified for centre and primary or secondary care setting. The primary outcome was improvement in Montgomery-Åsberg Depression Rating Scale (MADRS) score five weeks after randomisation using intention-to-treat analysis. Secondary outcomes included change in MADRS in patients with confirmed adherence to metyrapone versus those randomised to placebo, effects of metyrapone treatment on anxiety and quality of life measures, the persistence of treatment effects up to 6 months and the safety and tolerability of metyrapone. Trial registration ISRCTN45338259.

Findings

From March 2011 to December 2012, 165 patients were recruited with 86.7% reaching primary outcome time point and 63.0% the final 6 months' assessment. At 5 weeks

there was no significant difference between groups on the primary outcome measure, MADRS improvement (estimated mean difference -0.51, 95% CI: -3.48, 2.46; p>0.1), or on any secondary outcome measures including Beck Depression Inventory, Clinical Anxiety Scale, State-Trait Anxiety Inventory or EQ-5D-3L Quality of Life. Similarly there were no significant differences in outcomes at any other time points. Restricting analysis to metyrapone treated patients with endocrinological evidence of medication adherence did not alter the findings. Metyrapone was well tolerated.

Interpretation

Metyrapone augmentation of antidepressants is not efficacious in a broadly representative population of patients with TRD within the UK National Health Service (NHS).

Funding

Efficacy and Mechanism Evaluation (EME) programme, which is a UK Medical Research Council and National Institute for Health Research partnership.

Introduction

Clinical guidelines recommend the use of antidepressant medication for the treatment of moderate to severe major depressive disorder (MDD).[1;2] However a significant proportion of patients fail to obtain an optimal outcome to both first and second line treatments, commonly described as treatment resistant depression (TRD).[3;4] Hypothalamic-Pituitary-Adrenal (HPA) axis abnormalities are often demonstrated in patients with mood disorders and there is increasing evidence that such dysregulation is associated with poor prognosis, defined both by non-response to antidepressants and by an increased likelihood of future relapse.[5;6] A Cochrane review suggested efficacy of anti-glucocorticoid augmentation of antidepressants in patients with MDD, with the largest effect size seen with metyrapone,[7] a cortisol synthesis inhibitor. The data relating to metyrapone was drawn from a single positive double-blind, placebo controlled randomised study of metyrapone (250mg four times daily for 3 weeks) in a small sample of 63 depressed inpatients in Germany.[8]

The primary aim of the "Antiglucocorticoid augmentation of anti-Depressants in Depression" (ADD) study was to test the proof-of-concept of metyrapone previously established by Jahn and colleagues[8] in the most clinically relevant population. The ADD study examined the efficacy, tolerability and safety of metyrapone (500mg twice a day) for 21 days in augmentation of conventional serotonergic antidepressants in a randomised placebo controlled trial (RCT) in patients with MDD who had not responded to at least two courses of antidepressants in their current episode (i.e. patients with TRD). This is the stage in treatment sequencing at which the current UK National Institute for Health and Clinical Excellence (NICE) guidelines for the management of depression[1] recommend antidepressant augmentation strategies

following failure of antidepressant monotherapies. Given that few patients with depression are currently admitted to inpatient psychiatric units in the UK,[9] the study recruited a primarily outpatient (primary and secondary care) UK National Health Service (NHS) population. To date, all published studies of the use of anti-glucocorticoids in patients with TRD have used short treatment periods of 1-3 weeks,[7] which can appear counter-intuitive in such a potentially chronic condition. However, evidence suggests that the clinical effects of antiglucocorticoids on HPA axis function persist after their administration has ceased.[10;11] The persistence of effects on depressive symptoms and quality of life was therefore examined in the ADD study at 21 weeks after stopping metyrapone treatment compared with only 2 weeks' follow-up in the earlier study.[8] The hypotheses being tested and the study protocol are published elsewhere.[12]

Methods

Study Design and Participants

The ADD study was a multicentre two-arm (1:1 allocation), parallel group, double-blind, randomised, placebo-controlled superiority trial of augmentation of serotonergic antidepressants with metyrapone in patients with moderate to severe MDD who had failed to respond to adequate trials of at least two antidepressants in their current episode.

The study was conducted in 3 centres recruiting patients from 7 UK NHS Mental Health Trusts, and co-localised primary care services, in 3 areas: North East England (Northumberland Tyne and Wear NHS Foundation Trust and Tees, Esk and Wear

Valleys NHS Foundation Trust); North West England (Manchester Mental Health and Social Care NHS Trust, Greater Manchester West Mental Health NHS Foundation Trust and Pennine Care NHS Foundation Trust); and the Leeds/Bradford area (Leeds and York Partnership NHS Foundation Trust and Bradford District Care NHS Foundation Trust). Eligible patients were those aged 18-65 who had a Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV)[13] defined major depressive episode assessed using the Structured Clinical Interview for DSM (SCID) research version.[14] They were required to have a Hamilton Depression Rating Scale 17 item (HDRS17)[15] score of ≥18 at both week -2 and 0 relative to randomisation and commencement of experimental medication, determined using the GRID-HAMD for improved reliability[16], a Massachusetts General Hospital Treatment Resistant Depression (MGH-TRD) staging score of 2-10 as a measure of treatment resistance[17] and to be on a single agent or combination antidepressant treatment which included a serotonergic drug (a selective serotonin reuptake inhibitor, a tertiary amine tricyclic, venlafaxine, duloxetine or mirtazapine). At the point of randomisation, patients had to have been taking their current antidepressant medication, at the same dose, for a minimum of four weeks. Exclusion criteria included: another DSM-IV axis I diagnosis (this was later relaxed due to initial slow recruitment to allow patients with an anxiety disorder considered to be secondary to a primary diagnosis of depression); physical co-morbidity that would make metyrapone inappropriate, including untreated hypothyroidism, disorders of steroid production, cardiac failure, angina, myocardial infarction in the last 3 years and renal failure; pregnancy or breastfeeding; use of medication that would interact with metyrapone; dependence on alcohol or other drug(s) in the past 12 months, and/or current harmful use of such substances (defined

as meeting SCID criteria for harmful use or dependence); current or recent participation in a research study that could interfere with results.

A cohort of age and gender matched healthy volunteers (HVs) was recruited to act as comparators to the patient cohort in relation to HPA axis function (see below) as well as in a range of additional investigations including neurocognitive testing, magnetic resonance imaging and electroencephalography which will be reported elsewhere. HVs were recruited by advertisements in the University of Manchester and by emails sent to the Volunteer Database of the Institute of Neuroscience, Newcastle University. HVs were confirmed as having no current or past axis I disorder using DSM-IV criteria assessed using the SCID, had no first degree family history of mental illness, and had a HDRS17 score of less than 5. Exclusion criteria were the same as for the patient cohort.

Randomisation and blinding

Patients were randomised to metyrapone or placebo using permuted block randomisation, stratified by centre (North East, North West or Leeds/Bradford), level of care setting (primary or secondary care) and, for the North East and North West centres, by whether the patient agreed to participate in the mechanistic studies (described elsewhere[12]). The randomisation code was generated by an independent statistician in the Newcastle Clinical Trials Unit with the length of each block randomly set at 2 or 4 (with equal probability) unknown to study personnel to ensure concealment of allocation. Coded (numbered) packs of study drug and matched placebo were produced according to the randomisation schedule, by Catalent Pharma Solutions (http://www.catalent.com/index.php). Metyrapone capsules were over-

encapsulated (using Coni-Snap® capsules, Capsugel, Morristown, N.J, USA.) and appeared identical to the placebo capsules, and were dispensed from the clinical trials pharmacy at Northumberland, Tyne and Wear NHS Foundation Trust, Newcastle. Randomisation was through a centralised web-based system set up by the Newcastle Clinical Trials Unit with access to the randomisation code limited to the study pharmacist, independent statistician and database manager to ensure concealment of allocation. Treatment allocation remained blinded until after the last participant's final 24 week visit, apart from an independent examination of unblinded group data by the Data Monitoring and Ethics Committee (DMEC) which found no evidence of a relationship between those taking metyrapone and suicidal thoughts/attempts.

Procedures

Patient identification was supported by two hubs of the UK NIHR Mental Health Research Network (the North East and North West), a Comprehensive Local Research Network (West Yorkshire) and spanned primary and secondary care. A schedule of participant visits to study centres and assessments made on these occasions is detailed in the Supplementary Table 1 and described elsewhere[12], with all patients providing written informed consent.

HPA Axis Assessment

Saliva samples to determine the cortisol awakening response (CAR)[18] were obtained at week 0 (immediately prior to commencement of treatment augmentation) and then at weeks +3 (the day after cessation of experimental medication) and +5. Participants collected 5ml of saliva by passive drool[19] into a plastic collecting tube on wakening and at 15 minute intervals for a further hour. A further sample was collected by the

same method at 11pm the night before each of the three CAR assessments. The CAR data were analysed by calculating the area under the curve (AUC) of concentration against time, calculated using the trapezoidal method with respect to zero (AUCg) and with respect to the concentration on waking (AUCi) as previously described.[20]

In addition serum samples were taken at weeks -2 and +1 for analysis of cortisol precursors and metabolites. The increase in 11-deoxycortisol between weeks -2 and +1 was used as a measure of adherence to medication since this has been shown to be highly sensitive to treatment with metyrapone.[8;21]

Intervention

Participants continued their existing antidepressant regime and received study drug (metyrapone 500mg or placebo) twice daily, prescribed in the morning and at noon, for 21 days (matching previous studies[8]). All other treatments remained under the control of the patient's usual treating clinician, with encouragement to avoid medication changes between enrolment (week -2) and the primary outcome time point (week +5) unless there was a compelling clinical reason to do so. Any such changes in treatment did not lead to the patient being excluded from analysis.

Outcomes

The primary outcome measure of mood was the Montgomery-Åsberg Depression Rating Scale (MADRS[22]) which was assessed at week 0 and then at weeks +3 (end of active treatment period), +5 (primary outcome time point), +8, +16 and +24 from the date medication was started (+/- 2 days). Secondary outcomes included rates of response (defined as a 50% or greater reduction in MADRS score) and remission

(defined as MADRS ≤ 10) at week+5, the Clinical Anxiety Scale (CAS[23]), Beck Depression Inventory (BDI[24]), State Trait Anxiety Inventory (STAI[25]) and Young Mania Rating Scale (YMRS[26]). Quality of life was assessed using the self-completed EQ-5D-3L instrument (http://www.euroqol.org/)[27] and tolerability using the Toronto Side Effects Scale (TSES).[28] Safety assessments in case of metyrapone induced hypocortisolaemia included serum cortisol at week +1 and measurement of sitting and standing blood pressure and urea and electrolytes at weeks +1 and +5.

Statistical analysis

The study was powered to detect a difference with a standardised effect size of d=0.5 (c.f. that of 0.63 found by Jahn and colleagues[8]) between groups in the change in MADRS scores between baseline and five weeks post-randomisation, requiring 85 patients per group for 90% power with an alpha of 0.05 (to allow for 10% attrition during the trial, the original aim was to randomise 95 per group). Due to slow recruitment, and with the agreement of the funder and DMEC, the power requirement was reduced to 80% requiring a sample size of 63 per group (70 per group allowing for 10% attrition). A full detailed Statistical Analysis Plan (SAP) was agreed with the DMEC prior to study completion and breaking of the blind (see Supplementary Material).

The primary outcome was analysed as an intention to treat analysis of covariance of MADRS scores at +5 week, covaried for baseline MADRS; strata (centres and whether the patient originated in primary or secondary care) and treatment groups were included as fixed effects. The persistence of change in MADRS score was assessed using repeated measures analysis of variance using data from all time points. Secondary outcomes were examined using the same methods. Side effects were

assessed using the TSES and YMRS. The YMRS was analysed using the mixed models approach described above. For the TSES the relative risk of individual symptoms in the two groups was calculated. Pre-specified per-protocol analyses according to adherence with medication were carried out, based on measurements of 11-deoxycortisol prior to randomisation (week -2) and at week +1. A conservative criterion was used to judge adherence in the metyrapone group of both the week +1 11-deoxycortisol concentration, and the increase between week -2 and week +1, being greater than the mean plus three times the SD of the placebo group (or week +1 11-deoxycortisol concentrations more than the placebo mean plus four times the SD of the placebo group for subjects missing a week -2 concentration).

The salivary cortisol concentrations at 11pm and the AUCs were non-normally distributed and were natural logarithm transformed and compared using paired (for changes over time in patients) or non-paired (comparing patients and HVs) t-tests, with equal variance not assumed. In the event of data being missing for time points used in the AUC analysis, imputation was conducted by inserting the mean of the values immediately before and after the missing time point. Imputation was not used for either the first or last data point with subjects being excluded in such circumstances, as was the case for subjects with more than one missing data point.

Statistical analysis was conducted using IBM SPSS Statistics for Windows, Version 20.0 (IBM Corp, Armonk, NY, USA. 2011) and Stata, Version 12 (StataCorp, Texas, USA. 2011). The study was registered on 21/12/2009 (ISRCTN45338259) under the public title “Antiglucocorticoid augmentation of antiDepressants in Depression: the ADD study”. Clinical Trial Authorisation was given by the Medicines and Healthcare

products Regulatory Agency (MHRA: EudraCT: 2009-015165-31). Ethical approval was granted by the Sunderland Local Research Ethics Committee (REC Ref No. 10/H0904/9) on 22/04/2010.

Role of the funding source

The funder of the study had no role in study design, data collection, analysis, data interpretation, manuscript writing, or decision to submit for publication. The authors had full access to all the data. All authors had final responsibility for the decision to submit for publication.

Results

- Figure 1 near here –

The first patient was randomised at the beginning of March 2011 and the last patient in mid-December 2012. The last 24 week follow up was in late June 2013 following a 7 month extension due to initial slow recruitment. The flow of patients through the study is detailed in the CONSORT diagram in figure 1. Overall 877 potential patients were identified, 237 (27.0%) came from primary care, 320 (36.5%) from secondary care and 310 (35.3%) as self-referrals following media exposure of the study or seeing posters in GP surgeries (in 10 cases the source of the referral was not clear). Note that all self-referrals were currently engaged in treatment within the NHS and were subsequently categorised and stratified on the basis of the level of care they were receiving (i.e. primary or secondary care). Of these, 284 patients were assessed for eligibility by a formal face to face screening. Of the 111 who did not meet inclusion criteria, 10 did not meet the criteria for a major depressive episode, 52 had HDRS17

item scores < 18, 17 had axis 1 disorders other than anxiety, 9 were on an inappropriate antidepressant and 3 had MGH-TRD staging scores outside the range of 2-10, 18 had physical disorders which excluded them and 5 had other miscellaneous exclusion criteria (3 patients were excluded for more than one reason). Eight patients subsequently dropped out before randomisation (i.e. between weeks -2 and 0), resulting in 165 patients being randomised (82 to placebo and 83 to metyrapone), exceeding the revised target and providing an 84% power to detect the effects hypothesised in the sample size determinations. Of those randomised, 143 (86.7%) completed the primary outcome assessments at +5 weeks. Thirty nine dropped out between week +5 and week +24 so that 104 (63% of those randomised) completed the study.

Overall 42% of patients were recruited from primary care (the proportion differing between sites: Newcastle and Durham 69%; Leeds-Bradford 40%; Manchester 14%). Baseline data of demographic and clinical characteristics are shown in Table 1. The placebo and the metyrapone groups were well balanced on key demographic variables and clinical characteristics. In general there were very few missing data items and only a small number of missing values needed to be imputed (Supplementary Table 2).

- Table 1 & Table 2 near here -

Baseline MADRS scores by site and patient origin are shown in Table 2. Scores of patients recruited from secondary care were significantly higher than those from primary care (mean difference 3.4: 95% CI: 1.5, 5.3). Figure 2 shows the MADRS scores of patients in the two treatment groups (metyrapone and placebo) over the course of the entire study. Across the two treatment arms considered together there

was a significant reduction in MADRS scores of 6.0 points (95% CI: 4.5, 7.5) between baseline and week +5. However, there was no significant difference between treatment groups on the primary outcome (covarying for baseline severity, study centre and primary or secondary care location of patients) with an adjusted week +5 MADRS score difference of 0.51 points (95% CI: -3.48, 2.46). This lack of effect of metyrapone was consistent across analyses using all combinations of the three covariates – see Supplementary Table 3). Repeated measures analysis of variance using data from all assessment time points with the same covariates also found no significant difference between treatment arms. Consistent with this, neither response nor remission rates were significantly different between treatment groups (Table 3), with odds ratios (based on a logistic regression model adjusted for site and origin of patient; metyrapone/placebo) of 0.95 (95% CI: 0.41, 2.20) and 0.97 (95% CI: 0.40, 2.55) for response and remission respectively.

- Figure 2 and Table 3 near here –

Mean serum 11-deoxycortisol concentrations at week +1 were 0.9 nmol/L (SD=1.4) in the placebo group and 36.7 nmol/L (SD=65.9) in the metyrapone group, reflecting a 1.6 (SD=1.7) and 65.8 (SD=153.8) fold increase respectively compared to week 0. Three patients were missing week -2 11-deoxycortisol concentrations and hence had adherence judged on the basis of their week +1 concentrations. All others were judged on a combination of week +1 11-deoxycortisol plasma concentration and the increase between week -2 and week +1, as described in the methods section. Fifty two patients in the metyrapone group (75% of the potential sample) were deemed to be “adherent” and were compared with all 74 of the patients in the placebo group. This per-protocol analysis found no difference in week +5 MADRS scores co-varying for baseline

MADRS score and origin of patient, with the mean difference being -1.65 (95% CI: -4.94, 1.65). In line with this, there was no difference in response (23.1% vs 21.6%; adherent metyrapone vs placebo treated patients) or remission rates (17.3% vs 16.2%) at week +5 between the two treatment groups. The non-adherent metyrapone randomised patients showed a non-significant change in MADRS score from baseline to week +5 (-2.94, 95%CI: -6.82, 1.18), while both the adherent metyrapone and placebo randomised patients showed a significant reduction (-7.21, 95%CI: -9.60, -4.79; -5.77, 95%CI: -7.93, -3.66) respectively.

Of the 165 patients randomised, saliva samples were obtained from 151 (92%) for measurement of 11pm cortisol concentrations and CAR at week 0. A total of 67 HVs age, gender and IQ matched to the patients (see Table 1) were recruited. Saliva samples for 11pm cortisol concentrations and CAR data were available for 60 (90%). The cortisol data from both the patients and HVs is shown in Figure 3. The 11pm cortisol was greater in patients (2.48 nmol/L, SD=4.97) than HVs (1.38 nmol/L, SD=1.38; t=2.2, df=132.3, p=0.032 on transformed data). Neither AUCg (patients' untransformed mean=1.36, SD=1.06; HVs' mean=1.38, SD=0.70; t=0.72, df=209, p=0.481) nor AUCi (patients' mean=0.30, SD=0.73; HVs' mean=0.23, SD=0.74; t=0.64, df=209, p= 0.526) differed between patients and HVs.

- Figure 3 near here -

Baseline and week +3 11pm and AUC salivary cortisol data were available from 125 (11pm) and 123 (AUC) patients, 59 (47%, 11pm) and 57 (46%, AUC) of whom were randomised to metyrapone. In these latter patients, there was no difference between baseline and week +3 11pm cortisol concentration (2.48 nmol/L SD=3.31 vs 3.31

nmol/L SD=4.69; t=1.13, df=58, p=0.26), AUCg (mean=1.39, SD=1.2 vs mean=1.41, SD=1.07; t=0.10, df=56, p=0.92) or AUCi (mean=0.26, SD=0.86 vs mean=0.27, SD=0.65; t=0.99, df=56, p=0.33).

To assess the effect of HPA axis function on response to metyrapone, an analysis of the effect of metyrapone on week +5 MADRS scores was conducted covarying for baseline 11pm salivary cortisol concentration, AUCg or AUCi, or the difference in 11pm cortisol concentrations, AUCi or AUCg at week +3 compared to week 0. A lack of effect of metyrapone was confirmed in all analyses (for details see Supplementary Table 4).

Analysis of secondary outcomes similarly yielded estimated effects of metyrapone that were small and not statistically significant, both at the primary outcome time point of week +5, and using a repeated measures analysis of variance with all assessment time points (see supplementary figures 1-6. Data regarding all clinical outcome measures detailed in Supplementary Table 5).

Twelve serious adverse events were reported for 10 of the patients who were randomised to treatment (4 in the metyrapone group and 6 in the placebo group), none of which were judged to be related to medication. Most occurred well after the 3-week medication period and/or were related to pre-existing conditions (see Supplementary Table 6). A total of 229 adverse events were reported, of these 83 (38%; 95% CI: 32%, 45%) were thought to be possibly related to study medication and 18 (8%; 95% CI: 4%, 12%) were thought to be definitely related to study medication. Twelve (5%; 95% CI: 2%, 9%) of the events led to adjustment, interruption or discontinuation of study

medication. One hundred and five of the 165 patients (63.6%; 95% CI: 55.8%, 71.0%) had at least one adverse event with the median number of 1 (range: 0-13) and a mean of 2.0 (95% CI: 1.6, 2.4). The incidence rate ratios (risk in group randomised to metyrapone relative to risk in group randomised to placebo) were: unadjusted estimate = 1.34 (95% CI: 0.90, 1.98); estimate adjusted for centre and origin of patient = 1.41 (95% CI: 0.98, 2.03). Restricting the analysis to events classified as “possibly” or “definitely” related to study medication the corresponding estimates are: unadjusted estimate = 1.71 (95% CI: 0.98, 3.01); estimate adjusted for centre and origin of patient = 1.92 (95% CI: 1.14, 3.24). Similar findings were made with regards to the TSES. Across all 32 symptoms rated by this scale, there was no difference between the randomised groups in terms of incidence. There were two symptoms where the difference was significant at the 5% level (delayed ejaculation and weight loss both being greater in the placebo group). No differences were seen in the incidence of CNS side effects (for full details of all TSES rated side effects see Supplementary Table 7). There was no difference between treatment groups on the YMRS consistent with a lack of risk of mania following treatment with metyrapone in this patient population.

There were six instances of hypocortisolaemia during the study (one on placebo and 5 on metyrapone). In all cases these were asymptomatic based on a Standard Operating Procedure for PIs on the management of low cortisol that included information on associated symptoms. As part of a Standard Operating Procedure all patients with hypocortisolaemia had their lying and standing BP and urea and electrolytes checked but no abnormalities were detected in these measures. Medication was continued and repeat cortisol measurements at later dates returned to normal.

The DMEC requested that the risk of events indicating raised levels of suicidality be compared between the two trial arms. Using data from the suicide risk assessment tool (drawn from the Mini-International Neuropsychiatric Interview[29]) , the incidence rate ratio (metyrapone/placebo) based on a negative binomial regression model adjusting for centre and origin of patient care was 0.47 (95% CI: 0.17, 1.32) suggesting no evidence of an increased risk of events associated with suicidality in patients randomised to metyrapone. The estimated impact of metyrapone on suicidality score (as measured by item 10 on the MADRS) was a change of -0.15 (95% CI -0.53 and 0.24).

Patients randomised to metyrapone were less likely to attend follow-up visits than patents randomised to placebo (Figure 4). Fitting a Cox proportional hazards model the estimated hazard ratio was 0.57 (95% CI: 0.35, 0.93), however, most of the divergence occurred in the period after active treatment ended at week +3.

- Fig 4 near here -

Discussion

The key finding of the ADD study is that in a UK NHS population mainly of out-patients with TRD in primary or secondary care, augmentation of serotonergic antidepressants with metyrapone is ineffective. This result remained unchanged when only data from

patients with clear evidence of adherence to study medication were included in the analysis. A lack of effect was also seen on all secondary outcome measures.

The absence of a clinical response may arguably be consequent on the absence of a cortisol response to metyrapone. However in the smaller positive study by Jahn and colleagues, cortisol levels were similarly unaffected[8] and a previous study has also shown a lack of correlation between change in cortisol concentrations and improvement in mood with metyrapone.[30] The lack of change in cortisol following metyrapone treatment may be due to homeostatic mechanisms acting to maintain cortisol concentrations by increasing HPA axis drive. In the study by Jahn and colleagues, ACTH and 11-deoxycortisol concentrations were robustly increased after metyrapone treatment.[8] We did not measure ACTH, but 11-deoxycortisol concentrations were similarly affected by metyrapone in this study.

It can also be conjectured that the relatively normal baseline HPA axis function in our sample may have militated against a clinical response to metyrapone. Our group has previously shown that the extent of HPA axis dysregulation predicts clinical response to a different anti-glucocorticoid treatment in bipolar depression.[31] However, counter to this, translational studies from our group show that anti-glucocorticoid strategies engender an increase in the prefrontal cortex serotonin response to a selective serotonin reuptake inhibitor (SSRI), even in rats with normal HPA axis function.[32] In the current study, while patients had a small but significantly increased 11pm cortisol concentration compared to matched HVs, there was no difference in CAR AUCg or AUCi. These AUC measures have been argued to represent total cortisol output and the degree to which the HPA axis can activate[20] and hence reflect HPA axis function

better than a single salivary cortisol concentration. Either way, change in MADRS score in patients treated with metyrapone did not relate significantly to baseline HPA axis function, or to measures of change in HPA axis function following treatment with metyrapone. It is however important to note that our study was not sufficiently powered to show whether metyrapone is efficacious in the sub-sample of hypercortisolaemic TRD patients.

The ADD Study is the largest RCT of metyrapone augmentation for TRD conducted to date. A strength of the current study is that, given the broad inclusion criteria and minimal exclusion criteria, it is generalizable to the large numbers of patients in the NHS who have TRD. However, there were a large number (n=712; 81%) of exclusions between referral and randomisation so we cannot be completely confident that our sample is fully representative of the clinical condition in the community at large.

The degree of treatment resistance was assessed using the MGH-TRD staging scale[17] which included taking a treatment history and examination of hospital and GP records. We chose a minimum MGH-TRD score of 2 for inclusion which represents a failure to respond to at least 2 antidepressants, given usual UK practice in primary care of not augmenting or combining medications for depression until after this stage.[1] Beyond this point there is great divergence in practice in the treatment sequencing for individual patients, with patients being referred to secondary care at different stages by individual clinicians. The maximum MGH-TRD score for inclusion was set as 10. In practice this means 5-6 treatment trials, of at least minimum duration, of different antidepressants or strategies allowing for dose optimisation and augmentation/combination of drugs in some trials. Use of electroconvulsive therapy

(ECT) was not an exclusion criterion. However this scores 3 points on the MGH-TRD and given that most patients receiving ECT will also have had at least two antidepressants (often with dose optimisation and augmentation), few patients treated with ECT score under the maximum cut off of 10. Of the 165 patients randomised just 7 (4%) had received ECT.

This sample in which 45% of patients were from primary care and where the vast majority of the 55% from secondary care were outpatients, is very different from that in the exclusively inpatient study of metyrapone augmentation of conventional antidepressants carried out by Jahn and colleagues.[8] The degree of treatment resistance is not described for the patients included in the Jahn study. Given that recruitment was on the basis of the patients being acutely unwell and requiring admission, it may be that they were less treatment resistant than those in the ADD Study, where defined TRD was an inclusion criteria. In addition, the metyrapone group in the Jahn study had a slightly higher mean MADRS score at baseline of 31.5 (SD=7.6) compared with 28.1 (SD=5.5) in the current study. However, the reverse was true of baseline BDI scores of 30.0 (SD=8.4) in the Jahn et al. study compared with 35.6 (SD=10.9) in the current study. High BDI scores and a high BDI/MADRS ratio have been associated with poor outcome in TRD.[33] Our patients therefore had clinical characteristics which are associated with worse outcomes. This is consistent with the overall low response and remission rates seen in the current study. These factors may explain the differences in findings between those of Jahn et al. and those of the current study.

There are a number of limitations to consider in relation to the study. Although the trial did not reach its original target of 90% power, it achieved 84% power with regards to the primary outcome measure. In the binary outcomes of response and remission, the very wide confidence interval suggests that the study is not adequately powered to detect differences in these measures of outcome, although response and remission rates were almost identical in the two groups. The 95% confidence intervals of the differences in MADRS scores between groups indicate that it is not possible to exclude an advantage to metyrapone of 3.5 points (or a disadvantage of 2.5 points). Therefore, while this study does not support the efficacy of metyrapone it cannot exclude a small effect (3.5 points comprising an effect size of 0.3). However, the post-hoc analysis which included only those patients defined as adherent to treatment supports the interpretation that metyrapone lacks efficacy in this population. This analysis is based on week +1 endocrine data and so it is possible that some patients who were adherent to metyrapone up to that point did not remain so for the subsequent 2 weeks of treatment. No assessment of the extent of adherence in the placebo group was conducted. The attrition rate in the follow up phase was somewhat higher in the metyrapone group than in the placebo group but it is unlikely that this would have significantly impacted on the primary clinical outcome measure since the difference in attrition was most marked after week +5. Nevertheless it is important to emphasise that the study shows no evidence of efficacy rather than necessarily evidence of no efficacy.

The assumption made in the study was that clinical effects on depression would be detectable after only 3-weeks treatment with metyrapone. However, apart from the Jahn et al study,[8] there is no empirical evidence to confirm this and it is not possible

to exclude that longer treatment with metyrapone could have had a positive effect. However current evidence suggests that antidepressant treatments start to work during the first week of treatment and separation between active and placebo arms is seen early even though statistical significance may not occur for a number of weeks.[34] There was no indication of this early effect occurring in our study (Figure 3), indeed there was less improvement on metyrapone at the end of active treatment, making it unlikely that extending treatment would have shown benefit.

The main conclusion is that in the population of depressed patients studied, the addition of metyrapone to standard serotonergic antidepressants is not effective and therefore cannot be recommended as a treatment option for TRD. A question remains as to why the ADD study was negative when an effect of anti-glucocorticoid treatment is supported by preclinical data[32] and by a previous RCT of metyrapone augmentation.[8] This may relate to the nature of the patients studied as discussed above and/or their relative lack of HPA axis dysfunction. Chronic depression has been shown to be associated with normal HPA axis function.[35] The initial hypercortisolaemia of depression may normalise with time in patients who continue to demonstrate symptoms; hence normal cortisol levels may be a consequence or a cause of chronic treatment resistant depression. This is an important issue for future research to clarify. Additionally, despite significant heterogeneity between different patient subgroups, there remains evidence for HPA axis dysfunction in depression.[36] HPA axis genes also appear to be central to the gene-environment interactions linking well-established aetiological factors, such as early life trauma, and the development of depression[37], as well as the genetic and epigenetic factors underlying the stress-diathesis model of depression.[38] Therefore there continues to be a need for further exploration of

treatments targeting the HPA axis for the prevention and treatment of depression. For example, it would be of interest to explore the efficacy of metyrapone augmentation in acutely depressed patients, particularly those with demonstrable hypercortisolaemia.

References

(1) National Institute for Health and Clinical Excellence. Depression: the treatment and management of depression in adults (update). 2009. Report No.: NICE Clinical Guideline 91.

(2) Anderson IM, Ferrier IN, Baldwin RC, Cowen PJ, Howard L, Lewis G, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. J Psychopharm 2008 Jun;22(4):343-96.

(3) Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006 Jan;163(1):28-40.

(4) Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-17.

(5) Young EA, Altemus M, Lopez JF, Kocsis JH, Schatzberg AF, DeBattista C, et al. HPA axis activation in major depression and response to fluoxetine: a pilot study. Psychoneuroendocrin 2004 Oct;29(9):1198-204.

(6) Vreeburg SA, Hoogendijk WJ, DeRijk RH, van DR, Smit JH, Zitman FG, et al. Salivary cortisol levels and the 2-year course of depressive and anxiety disorders. Psychoneuroendocrin 2013 Jan 10.

(7) Gallagher P, Malik N, Newham J, Young AH, Ferrier IN, Mackin P. Antiglucocorticoid treatments for mood disorders. Cochrane Database of Systematic Reviews 2008;(1):CD005168.

(8) Jahn H, Schick M, Kiefer F, Kellner M, Yassouridis A, Wiedemann K. Metyrapone as additive treatment in major depression: a double-blind and placebo-controlled trial. Arch Gen Psychiatry 2004 Dec;61(12):1235-44.

(9) Green BH, Griffiths EC. Hospital admission and community treatment of mental disorders in England from 1998 to 2012. Gen Hosp Psychiatry 2014 Jul;36(4):442-8.

(10) Rotllant D, Armario A. A single dose of metyrapone caused long-term dysregulation of the hypothalamic-pituitary-adrenal axis in the rat. Neurosci 2005;130(2):427-34.

(11) Gallagher P, Watson S, Elizabeth DC, Young AH, Ferrier IN. Persistent effects of mifepristone (RU-486) on cortisol levels in bipolar disorder and schizophrenia. J Psychiat Res 2008 Oct;42(12):1037-41.

(12) McAllister-Williams RH, Smith E, Anderson IM, Barnes J, Gallagher P, Grunze HC, et al. Study protocol for the randomised controlled trial: Antiglucocorticoid augmentation of anti-Depressants in Depression (The ADD Study). BMC Psychiatry 2013 Aug 3;13(1):205.

(13) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV). Washington, DC: American Psychiatric Association; 1994.

(14) Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry 1992 Aug;49(8):624-9.

(15) Hamilton M. Development of a rating scale for primary depressive illness. British Journal of Social & Clinical Psychology 1967 Dec;6(4):278-96.

(16) Williams JB, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, et al. The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. Int Clin Psychopharmacol 2008 May;23(3):120-9.

(17) Fava M. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry 2003 Apr 15;53(8):649-59.

(18) Wust S, Federenko I, Hellhammer DH, Kirschbaum C. Genetic factors, perceived chronic stress, and the free cortisol response to awakening. Psychoneuroendocrin 2000 Oct;25(7):707-20.

(19) Gallagher P, Leitch MM, Massey AE, McAllister-Williams RH, Young AH. Assessing cortisol and dehydroepiandrosterone (DHEA) in saliva: effects of collection method. J Psychopharmacol 2006 Sep;20(5):643-9.

(20) Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrin 2003 Oct;28(7):916-31.

(21) Otte C, Lenoci M, Metzler T, Yehuda R, Marmar CR, Neylan TC. Effects of metyrapone on hypothalamic-pituitary-adrenal axis and sleep in women with post-traumatic stress disorder. Biol Psychiatry 2007 Apr 15;61(8):952-6.

(22) Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Brit J Psychiatry 1979 Apr;134:382-9.

(23) Snaith RP, Baugh SJ, Clayden AD, Husain A, Sipple MA. The Clinical Anxiety Scale: an instrument derived from the Hamilton Anxiety Scale. Br J Psychiatry 1982 Nov;141:518-23.

(24) Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561-71.

(25) Kendall PC, Finch AJ, Jr., Auerbach SM, Hooke JF, Mikulka PJ. The State-Trait Anxiety Inventory: a systematic evaluation. J Consult Clin Psychol 1976 Jun;44(3):406-12.

(26) Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-35.

(27) EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208.

(28) Vanderkooy JD, Kennedy SH, Bagby RM. Antidepressant side effects in depression patients treated in a naturalistic setting: a study of bupropion, moclobemide, paroxetine, sertraline, and venlafaxine. Can J Psychiatry 2002 Mar;47(2):174-80.

(29) Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33.

(30) Raven PW, O'Dwyer AM, Taylor NF, Checkley SA. The relationship between the effects of metyrapone treatment on depressed mood and urinary steroid profiles. Psychoneuroendocrin 1996 Apr;21(3):277-86.

(31) Watson S, Gallagher P, Porter RJ, Smith MS, Herron LJ, Bulmer S, et al. A randomized trial to examine the effect of mifepristone on neuropsychological performance and mood in patients with bipolar depression. Biol Psychiatry 2012 Dec 1;72(11):943-9.

(32) Johnson DA, Grant EJ, Ingram CD, Gartside SE. Glucocorticoid receptor antagonists hasten and augment neurochemical responses to a selective serotonin reuptake inhibitor antidepressant. Biol Psychiatry 2007 Dec 1;62(11):1228-35.

(33) Rane LJ, Fekadu A, Wooderson S, Poon L, Markopoulou K, Cleare AJ. Discrepancy between subjective and objective severity in treatment-resistant depression: prediction of treatment outcome. J Psychiatr Res 2010 Nov;44(15):1082-7.

(34) Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? A meta-analysis. J Clin Psychiatry 2005 Feb;66(2):148-58.

(35) Watson S, Gallagher P, Del Estal D, Hearn A, Ferrier IN, Young AH. Hypothalamic-pituitary-adrenal axis function in patients with chronic depression. Psychol Med 2002 Aug;32(6):1021-8.

(36) Stetler C, Miller GE. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. Psychosom Med 2011 Feb;73(2):114-26.

(37) Hornung OP, Heim CM. Gene-environment interactions and intermediate phenotypes: early trauma and depression. Front Endocrinol (Lausanne) 2014;5:14.

(38) Booij L, Wang D, Levesque ML, Tremblay RE, Szyf M. Looking beyond the DNA sequence: the relevance of DNA methylation processes for the stress-diathesis model of depression. Philos Trans R Soc Lond B Biol Sci 2013;368(1615):20120251.

Figure and Table Titles and Legends

Figure 1: Study CONSORT diagram.

Figure 2: Salivary cortisol concentrations of patients and matched healthy volunteers.

Data is plotted for salivary cortisol concentrations at 11pm and following awakening. The data following awakening is used to calculate the area under the curve (AUCg and AUCi) for the cortisol awakening response (CAR). Data represents means with standard error bars.

Figure 3: Montgomery Asberg Depression Rating Scale (MADRS) scores of patients randomised to metyrapone or placebo over time.

Data represents means with standard error bars.

Figure 4: Patient retention in study over time.

Kaplan-Meier survival curve for patients randomised to metyrapone (n=83) or placebo (n=82) remaining in the study.

Table 1: Patient and healthy volunteer demographic information.

Table 2: Patient baseline Montgomery Asberg Depression Rating Scale (MADRS) scores by site and source (primary versus secondary care).

Table 3: Response and remission rates at week +5 for patients randomised to metyrapone or placebo.

Acknowledgements

This project was supported by the Efficacy and Mechanism Evaluation (EME) Programme (ref number 08/43/39), funded by the Medical Research Council (MRC) and managed by the National Institute for Health Research (NIHR) on behalf of the MRC-NIHR partnership. Further support was obtained from Northumberland, Tyne and Wear NHS Foundation Trust. Neither funding source had any role in data collection, analysis or writing of the manuscript, or the decision to submit it for publication. The corresponding author had full access to all data and had final responsibility for submission of the manuscript.

We thank all of the trial participants and clinicians who assisted in their recruitment. We would also like to express thanks for the support from the North East and North West hubs of the Mental Health Research Network, Primary Care Research Networks and Comprehensive Local Research Networks and the Manchester Wellcome Trust Clinical Research Facility. We also thank the members of the external members of the Trial Steering Committee (chaired by Prof Philip Cowen) and the Data Monitoring and Ethics Committee (chaired by Dr David Christmas) for their support and guidance.

The views expressed in this publication are those of the author(s) and not necessarily those of the MRC, NHS, NIHR or the Department of Health.

Declaration of Interests

As described in the acknowledgements, this study was funded by the Efficacy and Mechanism Evaluation (EME) Programme. INF was the chief investigator on the grant application and RHMW, IMA,HCRG, PMH, TH, AJL, EM, SP, NS, JW, FHW and SW were co-applicants. None of the authors, or members of the ADD study team, have any financial relationship with companies that manufacture or sell metyrapone. Several of the authors have, or had, financial relationships with pharmaceutical companies manufacturing or selling psychotropics within the last 3 years, as declared to the journal. These include: HMW – Astra Zeneca, Bristol-Meyers-Squibb, Eli Lilly, Ferro Group, Janssen, Lundbeck, MyTomorrows, Otsuka, Pfizer, Roche, Servier, SPIMACO, Sunovian; IMA – Alkermes, Lundbeck, Lundbeck/Otsuka, Servier; HCRG - Bristol-Meyers-Squibb, Desitin, Gedeon Richter, Lundbeck, Otsuka, Hoffman LaRoche; PMH - Bristol-Meyers-Squibb, Eli Lilly, Janssen, Lundbeck, Otsuka, Roche, Servier, Sunovian, Takeda, Teva, Quantum Pharmaceuticals; SP – Merck Serono, Viropharma; BNPS - Bristol-Meyers-Squibb, Eli Lilly, Pfizer. All authors not listed reported that they had no declaration of interest in relation to this paper.

Authors contributions

All authors reviewed, revised and approved the final version of the manuscript.

R. Hamish McAllister-Williams was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in design of the study particularly reviewing its EEG aspects and recruitment. He wrote the first draft of the paper and incorporated comments from all other authors into the final manuscript.

Ian M. Anderson was co-applicant for the funding for the study, a PI for the Manchester site and was involved in design of the study and recruitment.

Andreas Finkelmeyer was a senior research associate for the study and analysed the HPA axis data reported in the paper.

Peter Gallagher conducted the review of all antiglucocorticoid treatments prior to the study and was involved in its conduct throughout.

Heinz C.R. Grunze was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in design of the study and recruitment.

Peter M. Haddad was co-applicant for the funding for the study, a PI for the Manchester site and was involved in design of the study and recruitment.

Tom Hughes was co-applicant for the funding for the study, a PI for the Leeds site and was involved in design of the study and recruitment.

Adrian J. Lloyd was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in the design of the imaging components of the study.

Chrysovalanto Mamasoula conducted all the statistics for the study.

Elaine McColl was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in the design of the study and its protocols and governance.

Simon Pearce was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in the design and supervision of the endocrine components of the study.

Najma Siddiqi was a PI for the Bradford site and was involved in the design of the study and recruitment.

Baxi N.P. Sinha was a PI for the Durham/Teesside site and was involved in the design of the study and recruitment.

Nick Steen was co-applicant for the funding for the study and designed and oversaw all the statistics for the study.

June Wainwright was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in design of the study and recruitment and liaison with the MHRN.

Fiona H. Winter was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in design of the study and recruitment.

Nicol Ferrier was the lead applicant for the funding for the study and the Chief Investigator.

Stuart Watson was co-applicant for the funding for the study, a PI for the Newcastle site and was involved in design of the study particularly reviewing its endocrine aspects and recruitment.

The ADD study team were involved in a range of activities, primarily recruitment, data collection and data entry.

Antidepressant augmentation with metyrapone for treatment resistant depression (The ADD Study): a double-blind, randomised, placebo controlled trial

McAllister-Williams et al.

Research in context

Evidence before this study

There is abundant evidence of abnormalities in hypothalamic-pituitary-adrenal (HPA) axis function in at least a proportion of patients with major depression disorder (MDD), with some patient cohorts demonstrating hypercortisolaemia. As a result a number of anti-glucocorticoid treatments have been investigated in patients with MDD, particularly those with treatment resistant depression (TRD). One such anti-glucocorticoid treatment is metyrapone. To date, one small (n=63) double blind randomised controlled trial of metyrapone augmentation of serotonergic antidepressants in German in-patients has been described (Search using PubMed with terms "depression" and "metyrapone"; May 2015). This found a positive difference 2 weeks after a 3 week course of metyrapone 1g daily, compared to placebo, with an effect size of 0.63 for the reduction in Montgomery-Asberg Depression Rating Scale (MADRS) score.

Added value of this study

The ADD study's aim was to examine if this demonstrated effect of metyrapone is evident in a predominantly out-patient population of patients with TRD recruited using broad inclusion criteria in a larger, more naturalistic, sample (n=165). Further the study aimed to examine tolerability and if any beneficial effect seen was sustained over the longer term (6 months). As such the study was intended to inform whether metyrapone augmentation was a realistic therapeutic option in every-day clinical practice.

Implications of all the available evidence

Contrary to the previous double-blind RCT of metyrapone augmentation in TRD, the ADD study failed to demonstrate any beneficial effect of metyrapone augmentation. Treatment, however, was well tolerated. The ADD study therefore suggests that metyrapone augmentation of serotonergic antidepressants is not an option for TRD in routine clinical practice at this time. Further research is required to clarify the extent and longitudinal course of HPA axis abnormalities in MDD and in TRD, and whether this may help identify individuals in whom anti-glucocorticoid treatments may be effective.

Fig 1

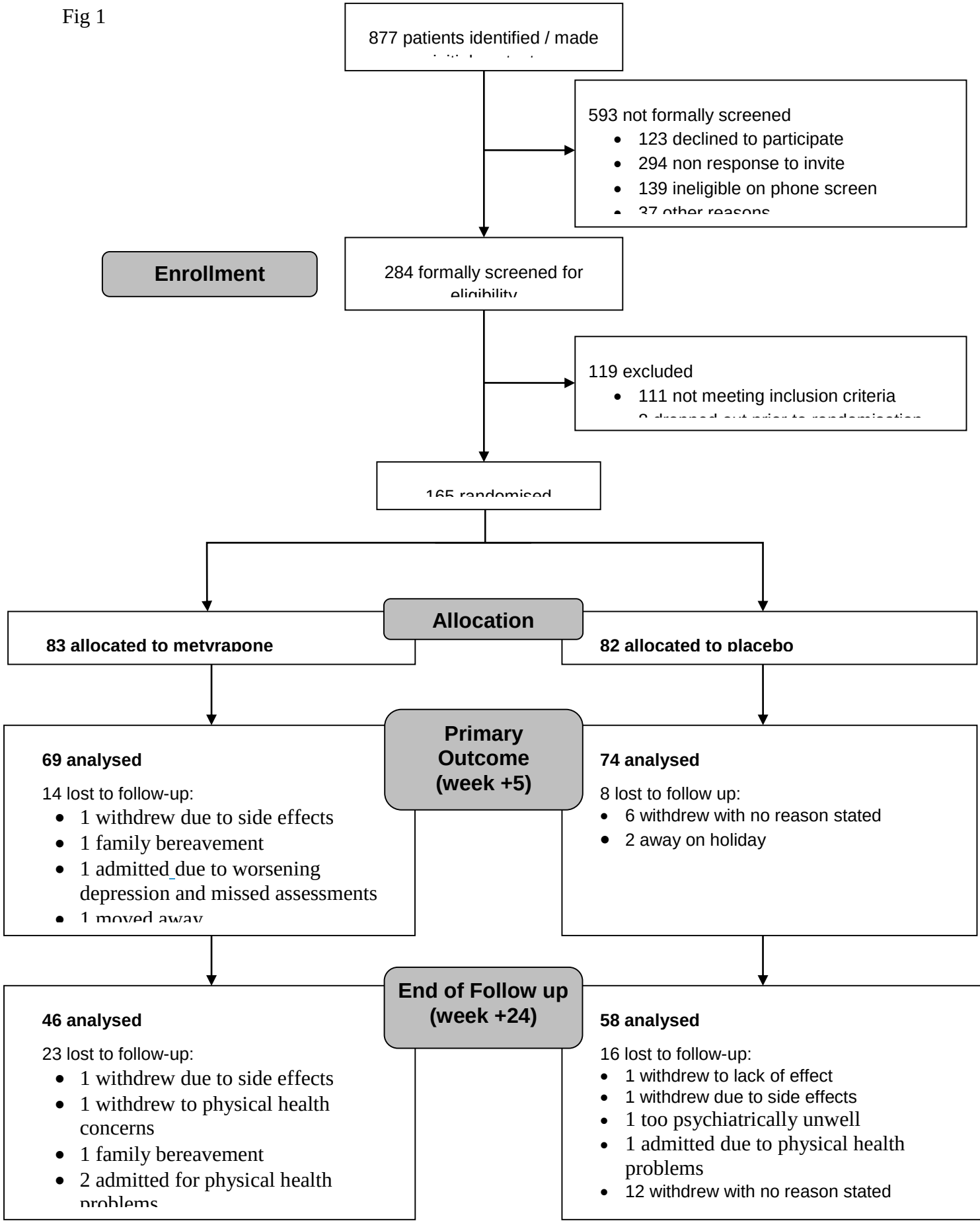

877 patients identified / made
593 not formally screened
123 declined to participate
294 non response to invite
139 ineligible on phone screen
Enrollment
284 formally screened for
119 excluded
111 not meeting inclusion criteria
165 randomised
Allocation
83 allocated to metyrapone
82 allocated to placebo
Primary Outcome (week +5)
69 analysed
14 lost to follow-up:
1 withdrew due to side effects
1 family bereavement
1 admitted due to worsening depression and missed assessments
74 analysed
8 lost to follow up:
6 withdrew with no reason stated
2 away on holiday
End of Follow up (week +24)
46 analysed
23 lost to follow-up:
1 withdrew due to side effects
1 withdrew to physical health concerns
1 family bereavement
2 admitted for physical health
58 analysed
16 lost to follow-up:
1 withdrew to lack of effect
1 withdrew due to side effects
1 too psychiatrically unwell
1 admitted due to physical health problems
12 withdrew with no reason stated

Fig 2

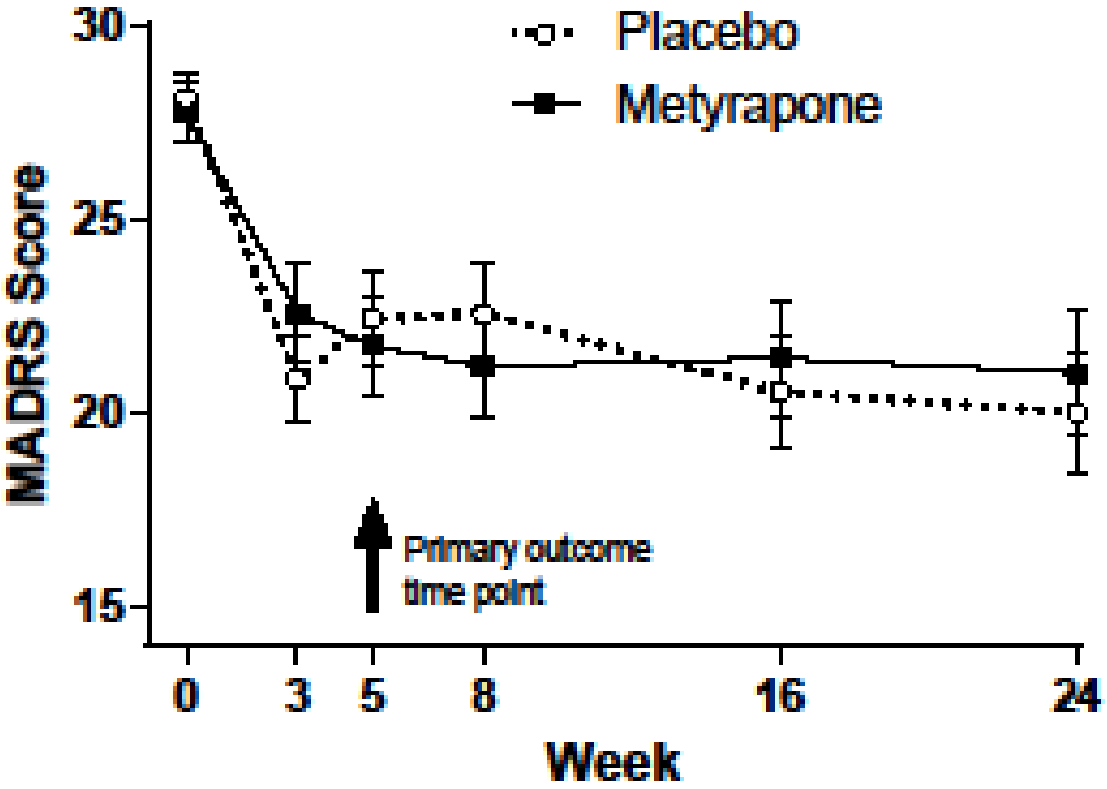

Placebo
Metyrapone
MADRS Score
30
25
20
15
Primary outcome
time point
0
3
5
8
16
24
Week

Fig 3

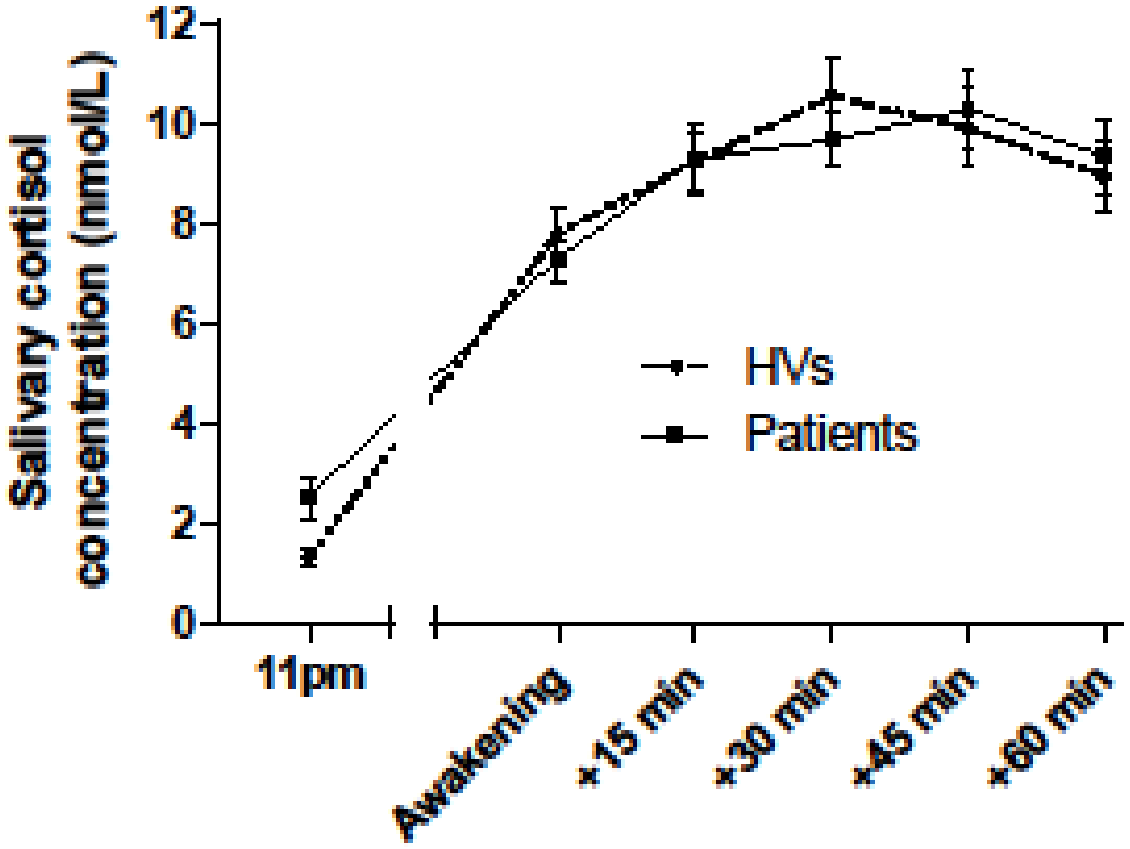
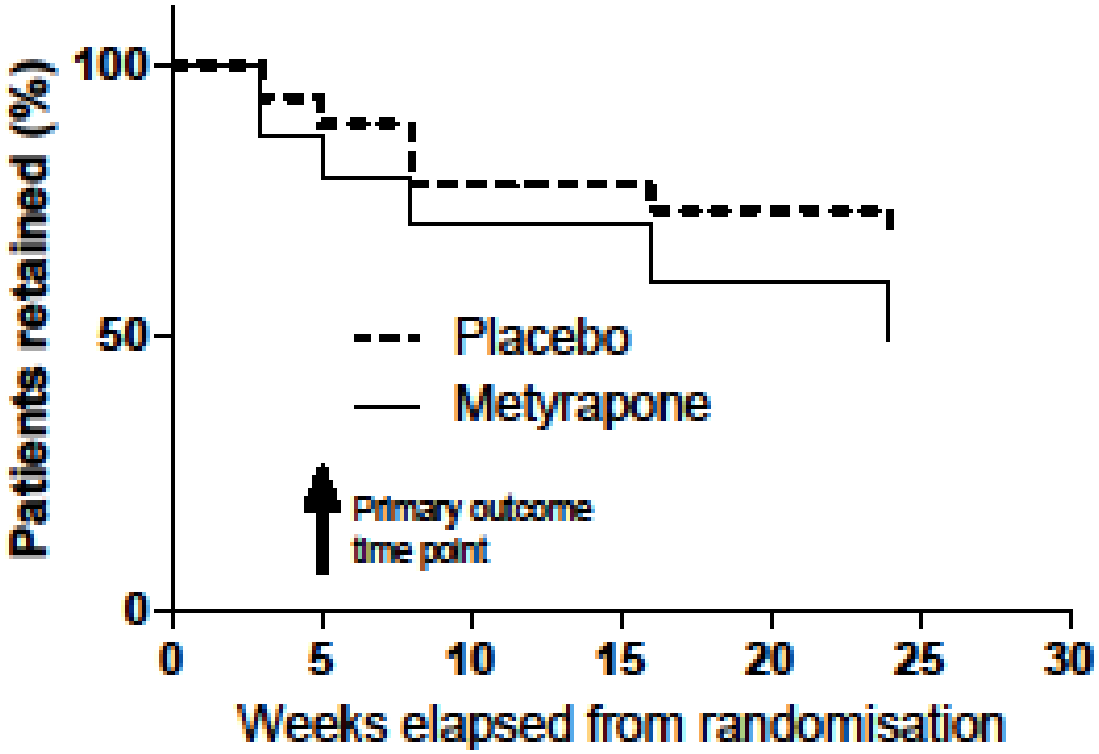

Salivary cortisol concentration (nmol/L)
12
10
8
6
4
2
0
HVs
Patients
11pm
Awakening
+15 min
+30 min
+45 min
+60 min

Fig 4


Patients retained (%)
100
50
0
Placebo
Metyrapone
Primary outcome
time point
0
5
10
15
20
25
30
Weeks elapsed from randomisation

Antidepressant augmentation with metyrapone for treatment resistant depression (The ADD Study): a double-blind, randomised, placebo controlled trial

McAllister-Williams RH et al.

Supplementary Appendix

The ADD Study Team: Hamid Alhaj[1,2], Tanefa Apekey[3], Jane Barnes[4], Sally Barker[2], Samantha Bulmer[1,2], Stephanie Clutterbuck[1], Julie Doughty[4], Claire Farrow[5], John Gray[1], John Hiley[6], Allan O. House[7], Brett Jones[1], Sophie Landa[1], Anna Massey[2], Ella Roelant[8], Faye Ryles[5], Paul Sigalas[1,2], Eleanor Smith[1,2], Chris Speed[4], Lucy Stevens[1], Lisa Svennson[1], Cathy Symonds[9], Helen Watkinson[2], Kate Williams[9], Sarah Yates[1,2].

Affiliations of the ADD Study Team:
1 - Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
2 – Northumberland Tyne and Wear NHS Foundation Trust, Newcastle upon Tyne, UK
3 - Greater Manchester West Mental Health NHS Foundation Trust
4 - Newcastle Clinical Trials Unit, Newcastle University, Newcastle upon Tyne, UK
5 – Tees, Esk and Wear Valleys NHS Foundation Trust, UK
6 – Bradford District Care NHS Foundation Trust, UK
7 - Institute of Health Sciences, University of Leeds, Leeds, UK
8 - Leeds and York Partnership NHS Foundation Trust, Leeds, UK
9 - Neuroscience and Psychiatry Unit, Manchester University, Manchester, UK

The ADD Study Statistical Analysis Plan (clinical outcomes only)

Study objectives

Primary Clinical Objective:

- The primary objective is to determine whether metyrapone (500mg twice a day) for 21 days is efficacious in augmenting conventional serotonergic antidepressants in TRD in a UK NHS primary and secondary care setting. This is to be assessed by Montgomery-Åsberg Depression Rating Scale (MADRS) scores measured two weeks post treatment (week +5 from randomisation), comparing patients treated with metyrapone to those treated with placebo.

Secondary Clinical Objectives:

1. To determine the clinical effect size at two weeks post-completion of treatment of a three-week course of metyrapone (vs placebo) augmentation of antidepressants in depressed patients who have failed to respond to at least two courses of antidepressants, in primary care and psychiatric outpatient clinics in the UK.
2. To assess whether the response is sustained for up to 21 weeks post cessation of Metyrapone.
3. To assess whether metyrapone augmentation improves patients' quality of life.
4. To assess the tolerability and safety of metyrapone augmentation in a large sample taken from a representative population of psychiatric outpatients and primary care patients with TRD.

Mechanistic Objectives:

NOTE: This version of the Statistical Analysis Plan only includes the mechanistic objectives related to the full study sample. Additional objectives related to sub-samples of patients that had additional investigations. These are referred to in the published protocol (McAllister-Williams et al. BMC Psychiatry 2013 Aug 3;13:205) and further details are available on request.

1. To assess whether metyrapone changes patients' HPA axis function.
2. To assess whether changes in HPA axis function are seen 2 weeks after metyrapone treatment.
3. To assess whether the change in HPA axis function correlates with clinical response.
4. To assess whether baseline HPA axis function predicts clinical response.
5. To assess if type and severity of childhood trauma (as assessed by the Childhood Trauma Questionnaire (CTQ) scores) predicts clinical response.

6. To assess whether the change in, or baseline assessment of, HPA axis function has specific genetic underpinning and whether this relates to, or explains better, the clinical response.

Study design:

Patient randomised double blind randomised controlled parallel trial with half the patients receiving Metyrapone and half the patients receiving a placebo.

Sample size

Two groups of 85 patients give 90% power to detect an effect size of 0.5 assuming a type 1 error rate of 5%. Allowing for 10% attrition during the trial, 95 per group would need to be randomised; 190 in total aged 18-65 years. 55 healthy control subjects, aged 18-55 with no personal or family history of mental illness are included.

The target number in the patient group was amended following a delay in starting recruitment and a slow early rate of recruitment, following agreement from the Data Monitoring and Ethics Committee (DMEC), the funder and the sponsor. A reduced power of 80% was agreed. Using the same assumptions as above, and allowing for 10% attrition prior to the primary outcome measure, the aim was to randomise 70 per group; 140 in total.

Study population

This is more fully defined in the ADD study protocol publication. Patients included had to meet criteria including i) having a DSM-IV confirmed diagnosis of a major depressive episode; ii) current symptoms of moderate to severe severity as defined by Hamilton Depression Rating Scale -17 item (HDRS17) score of $\geq$ 18 at week -2 and 0; iii) have TRD as defined by a Massachusetts General Hospital Treatment Resistant Depression (MGH-TRD) staging score of 2-10; iv) currently taking a serotonergic antidepressant.

CONSORT diagram

The flow of patients through the study will be described with the aid of a CONSORT diagram.

Characteristics of groups at baseline

The baseline characteristics of the study population will be summarised for each study group.

Compliance with medication and withdrawal from study medication

Adherence to medication will be assessed using measures of 11-deoxycortisol.

A 95% confidence interval for the relative odds of withdrawing from the two treatment arms will be reported.

Completeness of data

The level of missing data will be reported for each of the study measures of outcome.

General Analysis Considerations

Unless otherwise stated, all analyses will use two-tailed tests where appropriate with significance level set at 5%. Statistical packages used will include SPSS, STATA and R, as well as specialised packages for MRI/EEG analysis etc.

Primary outcome

The primary outcome will be assessed using an intention to treat analysis (patients allocated to the group to which they were randomised regardless of whether they received the treatment).

Primary analysis

- The estimated difference between the treatment groups at week 5 adjusting for randomisation strata will be estimated using a mixed model. The dependent variable will be MADRS score at week 5. The baseline score at week 0 will be included as a covariate. Differences between centres will be incorporated as a random effect; differences between primary and secondary care origin of the patients will be included as a fixed effect. (Statistical packages: SPSS, STATA, R).

Secondary analyses

- The difference between MADRS at week 5 and MADRS immediately before treatment (week 0) in two arms will be compared using an independent sample t-test utilising all patients in the study in an intention-to-treat analysis.
- The impact of missing data due to patients declining to participate in data collection will be assessed using joint modelling of "time to withdrawal from data collection" and MADRS scores.
- In addition to analysing the data using the conventional MADRS score, the above primary and secondary analyses will be repeated utilising the addition "atypical" depression items (rating hypersomnia and increased appetite). In this scoring of the MADRS, the highest score from the conventional sleep and atypical sleep items, and conventional appetite and atypical appetite items, will be used to calculate the total MADRS score.
- The intention to treat analyses above will be repeated in a "per protocol analysis" excluding patients who discontinued treatment during the three week treatment period. A second "per protocol" definition will utilise the 11-deoxycortisol measure of apparent concordance with treatment.

Secondary Clinical outcomes

1. In relation to clinical effect size:

a. Response (defined as 50% reduction in MADRS) and remission (defined as MADRS ≤ 10) rates at week 5 will be analysed using logistic regression procedures analogous to the mixed models described previously. Variation between centres will be incorporated as a random effect; differences between primary and secondary care origin of patients and the difference between those randomised to Metyrapone and those randomised to placebo will be included as fixed effects. Results will be presented in the form of a 95% confidence interval for the odds of response/remission.
b. In relation to CAS, BDI and STAI measures:
    i. The CAS, BDI and STAI will be analysed by fitting a sequence of nested models. The baseline model will include the score at week 5 as the dependent variable, the baseline score as a covariate and the difference between study groups as a fixed effect. Variation between centres will then be included as a random effect and the difference between primary and secondary care origin as a fixed effect. Nested models will be compared using a likelihood ratio test. The impact of compliance with treatment will be evaluated using the same approach (based on the measures of compliance previously defined). Results will be given in the form of unadjusted and adjusted 95% confidence intervals for the effect of treatment with Metyrapone.

2. In relation to whether the response is sustained:
    a. The persistence of change in MADRS, CAS, BDI and STAI scores will be assessed using repeated measures ANOVA, adjusting for randomisation strata of centre and primary vs secondary care, using all of the data points available.
    b. The speed of response and speed of remission (both defined using MADRS as described above) with Metyrapone vs. placebo will be assessed using survival analysis; time to response and time to remission will be analysed using a Cox proportional hazards regression model. Results will be given in the form of a 95% confidence interval for the hazard ratio between the two groups.
3. In relation to whether metyrapone improves patient's quality of life:
    a. The EQ-5D data will be analysed by fitting a sequence of nested models. The baseline model will include the score at week 5 as the dependent variable, the baseline score as a covariate and the difference between study groups as a fixed effect. Variation between centres will then be included as a random effect and the difference between primary and secondary care origin as a fixed effect. Nested models will be compared using a likelihood ratio test. The impact of compliance with treatment will be evaluated using the same approach (based on the measures of compliance previously defined). Results will be given in the form of unadjusted and adjusted 95% confidence intervals for the effect of treatment with Metyrapone.
4. In relation to how well metyrapone is tolerated:
    a. The Toronto Side Effects Scale scores will be analysed utilising data from all patients who had at least one dose of study medication. A second "per protocol" definition will utilise the 11-deoxycortisol measure of apparent concordance with treatment.
        i. The difference between TSES score at week 3 and TSES score immediately before treatment (week 0) in two arms will also be compared using an independent sample t-test utilising all patients in the study in an intention-to-treat analysis.
        ii. The change in TSES score over time will be compared using a repeated measures ANOVA including all available time points.
        iii. Symptoms described by patients will be grouped into clinically relevant clusters and analysed. This will include a specific focus on suicidal and any changes in this during or following treatment.
    b. The Young Mania Rating Scale (YMRS) will be analysed in a similar way to that described for the TSES.

Mechanistic Outcomes

In the full patient sample:

1. To assess whether metyrapone changes patients' HPA axis function:
    a. HPA function will be assessed by calculating the Area Under the Curve (AUC) for the CAR data, measuring the peak change in cortisol on waking by comparing the initial waking sample with the maximum, as well as examining the diurnal change in cortisol using the CAR and the 11pm saliva sample. The effect of metyrapone will be compared to placebo in relation to HPA axis function at week 3 (end of metyrapone treatment) using ANOVA covarying for the baseline at week 0.
2. To assess whether changes in HPA axis function are seen 2 weeks after metyrapone treatment:
    a. HPA axis function assessed as above comparing data at week 5 covarying for week 0.

3. To assess whether the change in HPA axis function correlates with clinical response:
    a. The correlation between change in HPA axis function (CAR AUC and maximal peak, and diurnal cortisol, both at week 3 and week 5 compared to week 0) with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared to week 0) will be examined.
4. To assess whether baseline HPA axis function predicts clinical response:
    a. The correlation between HPA axis function (CAR AUC and maximal peak, and diurnal cortisol) with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared to week 0) will be examined.
5. To assess if type and severity of childhood trauma (as assessed by the Childhood Trauma Questionnaire (CTQ) scores) predicts clinical response:
    a. The correlation between CTQ with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared to week 0) will be examined.
6. To assess whether the change in, or baseline assessment of, HPA axis function has specific genetic underpinning and whether this relates to, or explains better, the clinical response:
    a. The relationship between clinical response (as described above) and HPA axis function compared with a range of candidate polymorphisms will be examined.

Sub group analysis

The data will be analysed by dividing the sample into those with severe depression (defined as MADRS $\geq$30 at baseline week 0) or less severe depression.

Missing data

Scores for the various scales used will be calculated in accordance with their author's instructions. In the absence of any rule for missing data imputation will be used provided at least half the items in any scale have been completed (imputed missing value = mean value of non-missing items). Missing survival date will be dealt with by using the date of censoring and the status of the patient at that point. The secondary outcomes are being analysed using mixed models that make use of all data available. The analysis is adjusted to take into account that some subjects have provided more information than others. If there is a difference in survival rates we will have more "missing" outcome data in one group than another. If this situation occurs the data will be analysed using joint modelling as described above. Thus estimates of functional health status and quality of life will be adjusted for differences in survival rates. There will be no other data imputation.

Safety

The number of serious adverse events in each group will be reported along with a 95% confidence interval for the relative odds (between the two treatment groups) of at least one event being reported.

The number of non-serious adverse events will be compared using a negative binomial regression model. Results will be given in the form of a 95% confidence interval for the incidence rate ratio.

Implementation

An SAP Implementation Group will be formed to oversee the analysis of all ADD related data. The analysis of the clinical outcomes will be undertaken by the study team based in Newcastle. The mechanistic and cross sectional analyses will be undertaken by various individuals/centres with the agreement of the SAP Implementation Group.

Glossary of Abbreviations used in the Statistical Analysis Plan

5-HT	5-hydroxytryptophan, serotonin
11β -HSD	11β-hydroxysteroid dehydrogenase
ACTH	Adrenocorticotrophic hormone
AE	Adverse events
ANT	Attentional Network Test
AUCg	Area Under the Curve Ground
AUCi	Area Under the Curve Increase
BDI	Beck Depression Inventory
BFI-44	Big Five Inventory 44
BOLD	Blood-Oxygen Level Dependent
BSE	Between search errors
CANTAB	Cambridge Neuropsychological Test Automated Battery
CAR	Cortisol awakening response
CAS	Clinical Anxiety Scale
CBT	Cognitive behavioural therapy
CI	Chief Investigator
COM	Combined memory condition
CRF	Case report form
CTA	Clinical trial authorisation
CTQ	Childhood Trauma Questionnaire
CTIMP	Clinical trial of an investigational medicinal product
DHEA	Dehydroepiandrosterone
DMEC	Data Monitoring and Ethics Committee
DSM-IV	Diagnostic and Statistical Manual-IV
ECMT	Emotional Categorization and Memory Test
EEG	Electroencephalogram
EEM	Emotional enhancement of memory
EL	Emotional labelling
EM	Emotional matching
EME	Efficacy Mechanism and Evaluation
ERP	Event related potential
ESMT	Emotional Source Memory Task
EQ-5D	EuroQol Quality of Life scale
FEER	Facial Emotional Expression Recognition
FEP	Facial emotion processing
fMRI	Functional magnetic resonance imaging
FWE	Family Wise Error
GCP	Good Clinical Practice
GR	Glucocorticoid receptors
GRID-HAMD	GRID Hamilton Depression Rating Scale
HDRS17	Hamilton Depression Rating Scale – 17 item
HPA	Hypothalamic-pituitary-adrenal
HV	Healthy volunteers
HTA	Health technology assessment
IAPS	International Affective Picture Set
IMP	Investigational medicinal product
LDAEP	Loudness dependency of auditory evoked potentials
LTE	List of Life Threatening Experiences
LTP	Long term potentiation
MADRS	Montgomery-Asberg Depression Research Scale
MCV	Mean cell volume
MGH	Massachusetts General Hospital
MHRA	Medicines and Healthcare Regulatory Agency
MHRN	Mental Health Research Network
MR	Mineralocorticoid receptors
MRC	Medical Research Council

MTA	Material Transfer Agreement
NRES	National Research Ethics Service
NICE	National Institute of Clinical Excellence
NIHR	National Institute of Health Research
OLB	Object-location binding
OLM	Object-location memory
PCRN	Primary Care Research Network
phMRI	Pharmacological functional magnetic resonance imaging
PI	Principal Investigator
PIC	Participant Identification Centres
POM	Position-only memory
QOL	Quality of life
R & D	Research and Development
RAVLT	Rey Auditory Verbal Learning Test
REC	Research Ethics Committee
ROI	Region of Interest
RRS	Ruminative Responses Scale
RT	Reaction time
SAE	Serious adverse event
SARs	Serious adverse reactions
SM	Shape matching
SmPC	Summary of major product characteristics
SOP	Standard operating procedure
SCID	Structured Clinical Interview for DSM
SCQ	Social Circumstances Questionnaire
SPM8	Statistical Parametric Mapping
SSI	Site specific information
SSRI	Selective serotonin re-uptake inhibitor
STAI	State Trait Anxiety Inventory
SUSAR	Suspected unexpected serious adverse reaction
SWM	Spatial working memory
TSC	Trial Steering Committee
TSES	Toronto Side Effects Scale
VEP	Visual Evoked Potential
WSE	Within search errors
YMRS	Young Mania Rating Scale

Supplementary Table 1: ADD Study assessment schedule

	Enrolment	**Random -isation**	**Follow up**							
Time point	Week -2	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 8	Week 16	Week 24
Assessment of eligibility	√									
Informed Consent	√									
Assessment of baseline characteristics	√									
Experimental Intervention										
Assessment of depression severity – HDRS17[1]	√	√								
Assessment of Clinical symptoms – MADRS[2], CAS[3], BDI[4],STAI[5], YMRS[6]		√			√		√	√	√	√
Assessment of Quality of Life – EQ-5D		√			√		√	√	√	√
Assessment of side effects – TSES[7]		√			√		√	√		
Assessment of side effects and adverse events – self report		√	√	√	√	√	√	√	√	√
Suicide risk assessment	√	√	√		√		√	√	√	√
Pregnancy Test if indicated	√	√	√	√	√					
Assessment of concomitant medication	√	√	√	√	√	√	√	√	√	√
Measurement of HPA axis function (CAR plus 11pm saliva sample)		√			√		√			
Physical observations**	√***		√				√			
Blood Tests – U&E's, cortisol	√****		√				√			

** - Physical observations comprised sitting and standing pulse and blood, and pressure respiration rate.
*** - Screening physical observations also included height and weight

**** – Screening blood tests also including thyroid function tests, liver function tests and full blood count
[1]HDRS17 – Hamilton Depression Rating Scale – 17 item
[2]MADRS – Montgomery-Asberg Depression Rating Scale
[3]CAS – Clinical Anxiety Scale
[4]BDI – Beck Depressive Inventory
[5]STAI – State Trait Anxiety Inventory
[6]YMRS – Young Mania Rating Scale
[7]TSES – Toronto Side Effects Scale

Supplementary Table 2: Rates of missing and imputed data
A – Data at Baseline

		Number of participants at baseline for whom:-				
			Score has been calculated			
Measure	**n**	**Score is missing**	**All**	**Without imputation**	**With imputation**	**Mean number of imputed items***
MADRS	165	0	165	165	0	-
CAS	165	0	165	165	0	-
STAI - State	165	3	162	154	8	1.13
STAI - Trait	165	3	162	153	9	1.00
BDI	165	3	162	160	2	6.00
YMRS	165	1	164	164	0	-
EQ-5D	165	3	162	162	0	-

BDI = Beck Depressive Inventory; CAS = Clinical Anxiety Scale EQ-5D = Euroqual 5 dimensions; MADRS = Montgomery Asberg Depression Rating Scale; STAI = State Trait Anxiety Inventory

*The mean number of missing items for the subset of participants form whom a total score was calculated using imputation. In the absence of any rule for missing data imputation was used provided at least half the items in any scale had been completed. Imputed missing values were the mean value of non-missing items in that scale.

B – Data at primary end point

		Number of participants at primary endpoint for whom:-				Mean number of imputed items*
			Score has been calculated			
Measure	**n**	**Score is missing**	**All**	**Without imputation**	**With imputation**	
MADRS	143	0	143	143	0	-
CAS	143	0	143	143	0	-
STAI - State	143	4	139	136	3	2.00
STAI - Trait	143	4	139	137	2	1.00
BDI	143	4	139	135	4	3.75
YMRS	143	0	143	143	0	-
EQ-5D	143	4	139	139	0	-

BDI = Beck Depressive Inventory; CAS = Clinical Anxiety Scale EQ-5D = Euroqual 5 dimensions; MADRS = Montgomery Asberg Depression Rating Scale; STAI = State Trait Anxiety Inventory

*The mean number of missing items for the subset of participants form whom a total score was calculated using imputation (see Table 2A for details).

Supplementary Table 3: Effect of metyrapone on Montgomery Asberg Depression Rating Scale (MADRS) scores at week +5 covarying for illness severity, study centre and source of patients (primary versus secondary care) on Montgomery Asberg Depression Rating Scale (MADRS) scores at week +5

Included covariates when estimating the impact of metyrapone on MADRS scores five weeks post randomisation	**Regression coefficient**			**Significance**	
	B	**95% CI**		**t**	**p**
None[1]	-0.68	-4.23	2.87	-0.38	0.705
Baseline severity	-0.43	-3.45	2.63	-0.28	0.783
Baseline severity, primary v secondary care	-0.51	-3.55	2.53	-0.33	0.740
Baseline severity, centre[2]	-0.34	-3.39	2.70	-0.22	0.824
Baseline severity, primary v secondary care, centre[2]	-0.43	-3.47	2.61	-0.28	0.782
Baseline severity, primary v secondary care, centre[3][4]	-0.51	-3.48	2.46	-0.34	0.736
Baseline severity, centre[2], primary v secondary care, compliance with medication[5]	-1.65	-4.94	1.65	-0.99	0.33
[1] analysis of depression score five weeks post randomisation only [2] differences between centres fitted as fixed effects [3] differences between centres fitted as a random effect [4] pre-specified primary analysis [5] per protocol analysis (non-compliers omitted from analysis)					

Supplementary Table 4: Effect of metyrapone on Montgomery Asberg Depression Rating Scale (MADRS) scores at week +5 covarying for HPA axis variables.

Included covariates when estimating the impact of metyrapone on MADRS scores five weeks post randomisation	**Regression coefficient**			**Significance**	
	B	**95% CI**		**t**	**p**
Baseline 11pm cortisol	-0.70	-4.63	3.23	-0.35	0.725
Baseline CAR AUCg	-0.56	-4.37	3.25	-0.29	0.771
Baseline CAR AUCi	-0.56	-4.37	3.25	-0.29	0.771
Baseline to week +5 change in 11 pm cortisol	-1.51	-5.61	2.60	-0.73	0.468
Baseline to week +5 change in AUCg	-1.16	-5.22	2.90	-0.56	0.573
Baseline to week +5 change in AUCi	-1.13	-5.17	2.92	-0.55	0.582
NB – for comparison with the primary outcome covariate analysis see supplementary table 3. AUCg – Area under the curve with respect to zero; AUCi – Area under the curve with respect to cortisol concentration on waking; CAR – Cortisol awakening response					

Persistence of change in MADRS scores

For each individual, the MADRS score was recorded on up to five occasions following randomisation. These repeated measures were analysed using a mixed model in which we assume that for each patient the MADRS scores varied randomly about an individual mean score (with standard deviation σ_e) but that this mean varied randomly across patients (with standard deviation σ_u). A normal distribution was assumed for each of the random effects. Fitting an initial model with only a constant term (corresponding to the assumption that the mean MADRS score is the same at all five follow up visits) in addition to the random effects, the estimated mean MADRS score across the five follow up visits was 21.6 with 95% CI: 20.1, 23.0. There was significant variation of scores both within patients (σ_e= 6.57 with 95% CI: 6.17, 6.99) and between patients (σ_u= 8.58 with 95% CI: 7.54, 9.78).

Since we are interested primarily in changes in depression the second step was to include the baseline MADRS score as a covariate. This model indicated a significant reduction in depression from baseline at each of the five follow up visits. Baseline depression explained some of the variation between individual patients (the estimate of σ_u fell to 6.69) but there were still significant differences between patients.

Adding in differences between randomisation strata (sites and origin of patient-either primary or secondary care) did not explain very much additional variability. With the baseline score as a covariate there were no significant differences between these strata. The estimated difference in depression across all five follow up visits between patients on metyrapone and patients on placebo was 0.75 with 95% CI (-1.59, 3.10).

There was no evidence of any trend in depression over the follow up visits. Adding a linear trend over visits to the model the estimated mean change in MADRS score between consecutive visits was 0.27 (95% CI: 0.65, 0.10) and did not constitute a clinically important change in depression.

Supplementary Table 5: Means and standard deviations of all clinical outcome measures

		Metyrapone						**Placebo**					
		Base-line	Wk +3	Wk +5	Wk +8	Wk +16	Wk +24	Base-line	Wk +3	Wk +5	Wk +8	Wk +16	Wk +24
MADRS	Mean	27.7	22.6	21.71	21.2	21.4	21.0	28.1	20.8	22.4	22.6	20.5	20.0
	SD	6.7	10.9	10.9	10.4	11.0	11.1	5.5	9.9	10.6	10.8	11.5	11.6
	n	83	72	69	62	54	46	82	77	74	66	61	58
BDI	Mean	35.6	30.5	27.9	28.7	30.5	28.2	34.8	28.7	29.6	30.9	28	28.9
	SD	10.9	15.1	15.3	14.9	14.2	14.2	10.3	14	14.5	14.4	15.7	16.9
	n	81	69	67	61	52	47	81	75	72	63	61	57
CAS	Mean	9.5	8.5	8.5	8.7	8.5	7.9	10.0	7.4	8.2	8.5	7.7	7.4
	SD	4.5	6.2	6.0	6.0	5.0	5.7	4.6	5.2	5.8	5.6	4.9	5.3
	n	83	72	69	62	54	46	82	77	74	66	61	58
EQ-5D	Mean	0.37	0.44	0.50	0.47	0.49	0.49	0.37	0.46	0.47	0.45	0.45	0.44
	SD	0.30	0.33	0.35	0.34	0.33	0.35	0.31	0.33	0.35	0.36	0.36	0.37
	n	81	69	67	60	52	47	81	75	72	64	60	58
EQ-VAS	Mean	40.9	48.4	50.6	49.4	48.2	48.3	42.3	47.3	47.5	49.4	47	47.4
	SD	16.6	20.8	21.9	20.0	20.0	19.7	19.4	20.8	22.1	22.0	23.9	25.0
	n	80	65	63	59	51	47	79	74	71	64	61	58
YMRS	Mean	2.3	2.0	1.7	1.6	1.9	1.5	2.4	1.9	1.7	1.7	1.7	1.8
	SD	1.7	1.7	1.6	1.7	1.9	1.8	1.8	2.2	1.9	1.9	2.1	2.1
	n	82	72	69	62	54	46	82	77	74	66	61	59
STAI - State	Mean	42.8	42.9	42.1	41.3	42.3	42.4	41.1	39.9	40.3	40.7	41.2	41.3
	SD	6.5	7.1	6.9	7.2	6.1	6.1	5.8	5.5	6.0	7.0	7.3	6.2
	n	81	69	67	59	52	47	81	75	72	64	58	56

BDI – Beck Depressive Inventory; CAS – Clinical Anxiety Scale; EQ-5D – European Quality of Life-5 Dimensions; EQ-VAS – European Quality of Life Visual Analogue Scale; MDRS – Montgomery Asberg Rating Scale; SD – Standard Deviation; STAI – State Trait Anxiety Inventory; YMRS – Young Mania Rating Scale

Supplementary Table 6: Serious adverse effects

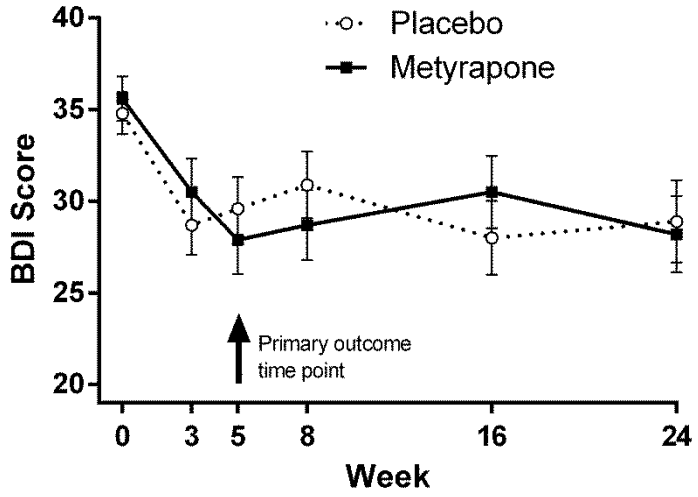
Event description	**Treatment allocation**
Breast cancer (diagnosis preceded randomisation), dehydration/leucopenia	Placebo
Groin infection/abscess	Placebo
Overdose (tramadol, pregabalin, amitryptiline and alcohol)	Metyrapone
Migraine with medication overuse headaches, right Holmes-Adie pupil	Metyrapone
Increased suicidal thoughts	Placebo
Dislocated knee/broken leg (bike accident)	Placebo
Overactive bladder	Metyrapone
Mood deterioration	Metyrapone
Patient took unknown amount of ketamine	Placebo
Increase in symptoms of depressed mood and re-emergence of suicidal ideation	Placebo

Supplementary Table 7: Side effect incidence on Toronto Side Effect Scale (TSES)

	Incidence							**Frequency**						**Severity**					
	Metyrapone		Placebo		Metyrapone/Placebo			Metyrapone			Placebo			Metyrapone			Placebo		
Side Effect	n	%	n	%	RR	95%CI		□	95%CI		□	95%CI		□	95%CI		□	95%CI	
Nervousness	53	77	53	72	1.1	0.9	1.3	2.9	2.5	3.3	2.5	2.2	2.8	2.6	2.3	2.9	2.4	2.1	2.7
Agitation	55	80	52	70	1.1	0.9	1.4	3.1	2.7	3.5	2.6	2.3	3.0	2.8	2.4	3.1	2.5	2.2	2.8
Tremor	28	41	21	28	1.4	0.9	2.3	1.9	1.5	2.2	1.6	1.4	1.9	1.7	1.4	2.0	1.5	1.2	1.7
Myoclonus	25	36	32	43	0.8	0.6	1.3	1.9	1.6	2.3	1.9	1.6	2.2	1.6	1.4	1.9	1.6	1.4	1.8
Abdominal pain	23	33	27	36	0.91	0.58	1.43	1.68	1.40	1.97	1.82	1.51	2.14	1.74	1.45	2.02	1.81	1.50	2.13
Dyspepsia	27	40	32	43	0.92	0.62	1.36	1.74	1.46	2.01	1.76	1.49	2.02	1.68	1.43	1.92	1.80	1.51	2.08
Nausea	27	39	29	39	1.00	0.66	1.50	1.68	1.43	1.94	1.62	1.39	1.85	1.81	1.51	2.11	1.66	1.42	1.91
Diarrhoea	21	30	26	35	0.87	0.54	1.39	1.52	1.29	1.75	1.59	1.35	1.84	1.54	1.31	1.76	1.62	1.36	1.88
Constipation	23	33	29	39	0.85	0.55	1.32	1.72	1.43	2.02	1.96	1.64	2.28	1.57	1.32	1.81	1.80	1.50	2.09
Decreased appetite	34	49	47	64	0.78	0.58	1.04	2.52	2.11	2.94	2.91	2.50	3.31	1.93	1.59	2.27	2.03	1.69	2.36
Increased appetite	16	30	19	26	0.90	0.51	1.61	1.48	1.22	1.73	1.50	1.26	1.74	1.29	1.10	1.48	1.38	1.15	1.60
Weakness or fatigue	56	80	61	82	1.0	0.8	1.2	3.6	3.2	4.0	3.6	3.3	4.0	3.1	2.8	3.5	3.5	3.1	3.8
Dizziness	24	53	25	34	1.0	0.7	1.6	1.7	1.4	2.0	1.7	1.4	1.9	1.7	1.4	2.0	1.7	1.4	1.9
Postural hypotension	25	36	32	43	0.8	0.6	1.3	1.6	1.4	1.8	2.0	1.7	2.3	1.6	1.4	1.90	1.7	1.4	2.0
Drowsiness	48	70	55	74	0.9	0.8	1.2	2.9	2.5	3.3	2.8	2.5	3.2	2.6	2.2	3.0	2.5	2.2	2.8
Increased sleep	15	22	18	24	0.9	0.5	1.6	1.5	1.2	1.8	1.6	1.3	1.9	1.3	1.1	1.6	1.4	1.2	1.7
Decreased sleep	46	67	55	74	0.90	0.72	1.11	3.09	2.67	3.51	3.09	2.71	3.47	2.90	2.48	3.31	2.85	2.48	3.22
Sweating	39	56	43	58	0.97	0.73	1.29	2.20	1.87	2.54	2.27	1.94	2.60	2.13	1.81	2.45	2.11	1.80	2.42
Flushing	31	45	29	39	1.15	0.78	1.69	2.00	1.67	2.33	1.73	1.47	1.99	1.81	1.52	2.10	1.70	1.42	1.98
Oedema	17	25	17	23	1.07	0.60	1.93	1.72	1.38	2.07	1.57	1.29	1.84	1.48	1.22	1.73	1.50	1.25	1.75
Headache	47	68	57	77	0.88	0.72	1.08	2.35	2.03	2.66	2.68	2.36	2.99	2.26	1.95	2.57	2.55	2.25	2.86
Blurred vision	23	33	22	30	1.12	0.69	1.82	1.67	1.39	1.95	1.62	1.35	1.89	1.54	1.30	1.77	1.49	1.27	1.70
Dry mouth	47	68	54	73	0.93	0.75	1.15	3.16	2.73	3.59	3.03	2.62	3.43	2.59	2.20	2.98	2.30	1.95	2.64
Anorgasmia	39	57	42	57	1.00	0.75	1.33	2.91	2.46	3.37	2.84	2.41	3.26	1.46	1.19	1.74	1.57	1.28	1.85
Increased libido	4	6	6	8	0.71	0.21	2.43	1.09	1.00	1.18	1.12	1.00	1.24	1.17	1.04	1.31	1.12	0.97	1.27
Decreased libido	33	48	37	50	0.96	0.68	1.34	2.71	2.25	3.17	2.66	2.23	3.09	1.84	1.50	2.18	1.74	1.43	2.05
Premat. ejaculation	1	3	1	3	1.00	0.07	15.28	1.13	0.87	1.39	1.03	0.97	1.10	1.13	0.87	1.39	1.00	-	-
Delayed ejaculation	1	3	8	27	0.12	0.02	0.91	1.13	0.87	1.39	1.60	1.18	2.02	1.13	0.87	1.39	1.48	1.08	1.88
Erectile dysfunction	9	29	8	27	1.09	0.48	2.45	2.13	1.47	2.79	1.93	1.31	2.55	1.39	0.95	1.83	1.43	1.02	1.85
Other, specify	0	0	0	0	-	-	-	1.00			1.00	-	-	1.00	-	-	1.00	-	-
Weight gain	20	29	15	20	1.43	0.80	2.56	1.64	1.34	1.93	1.43	1.20	1.67	1.51	1.24	1.77	1.43	1.18	1.68
Weight loss	7	10	18	25	0.41	0.18	0.92	1.14	1.03	1.26	1.47	1.24	1.69	1.07	0.97	1.18	1.04	0.99	1.09

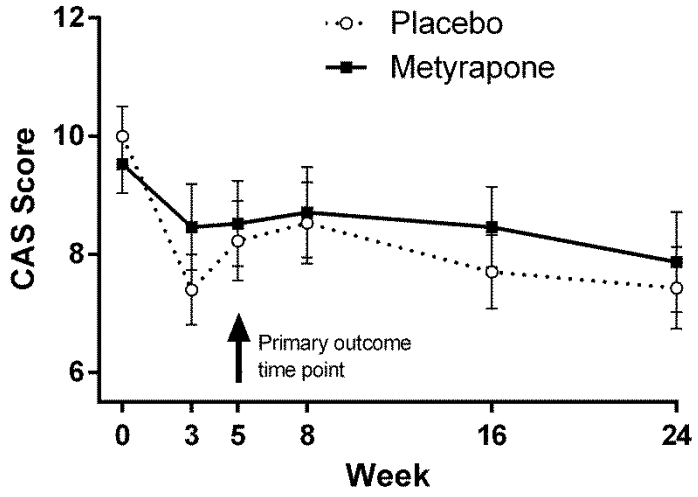
Effect of metyrapone on BDI scores 5 weeks post randomisation
As defined in the Statistical Analysis Plan, scores obtained five weeks post randomisation were analysed using analysis of covariance with baseline scores included as a covariate. This suggested a difference of -2.65 (95% CI: -6.53, 1.23) between the effects of metyrapone and placebo. Adjusting for random variation between centres and a difference between patients originating from primary and secondary care gave an adjusted estimate of -2.65 (95% CI: 6.41, 1.10) that was not significantly different from zero.

Supplementary Figure 1: Beck Depressive Inventory (BDI) scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.



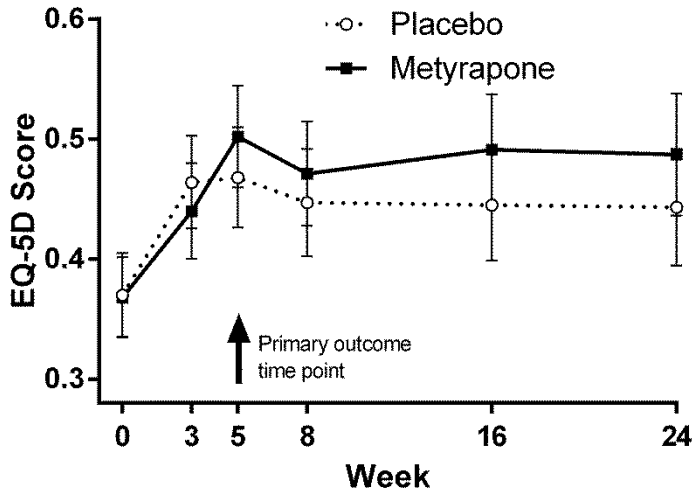
Effect of metyrapone on CAS scores 5 weeks post randomisation
Based on analysis of covariance of CAS scores recorded 5 weeks post randomisation (with baseline scores included as a covariate) the unadjusted estimate of the effect of metyrapone was 0.55 (95% CI: -1.21, 2.30). Adjusting for differences between centres and patient origin of care the estimated effect of metyrapone on CAS scores was 0.46, (95% CI: -1.20, 2.12). These estimates correspond to differences between groups that are neither statistically nor clinically significant.

Supplementary Figure 2: Clinical anxiety Scale (CAS) scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.



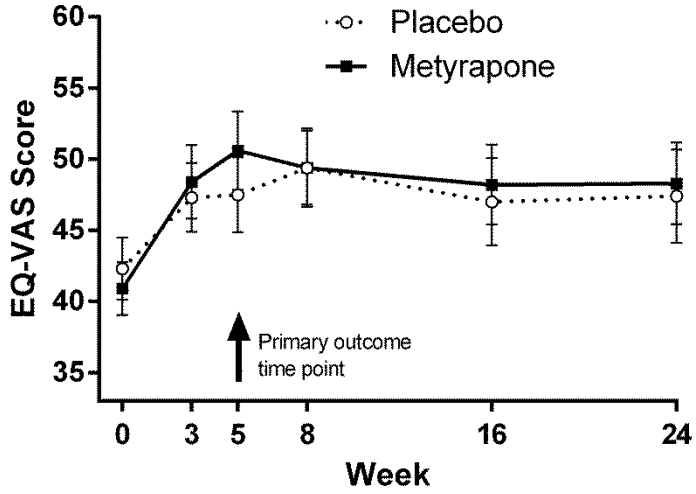
Effect of metyrapone on EQ-5D-3L health tariffs 5 weeks post randomisation
Based on a simple analysis of covariance model with baseline EQ-5D-3L value included as a covariate, the estimated effect of metyrapone was a change in EQ-5D-3L value of 0.014 (95% CI: -0.073, 0.101). Adjusting for variation between sites and origin of patient care the estimated effect of metyrapone was 0.015 (95% CI: -0.069, 0.099). There was no evidence of a beneficial effect of treatment with metyrapone on EQ-5D-3L.

Supplementary Figure 3: EuroQoL (EQ-5D-3L) scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.



Effect of metyrapone on EQ-VAS at 5 weeks post randomisation
The estimated effect of metyrapone at five weeks post randomisation (analysis of covariance with baseline scores included as a covariate) was 5.7 (95% CI: -0.8, 12.1). Adjusting for origin of patient and variation between centres the adjusted estimate is 5.6 (95% CI: -0.7, 12.0).

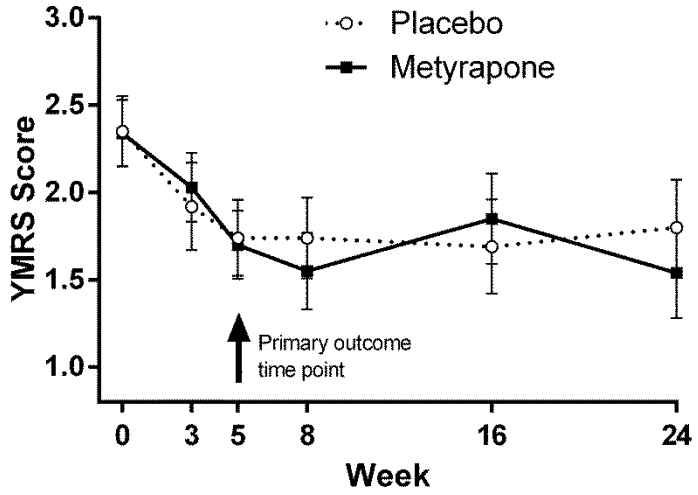
Supplementary Figure 4: EuroQoL visual analogue scale (EQ-VAS) scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.



The effect of metyrapone on YMRS scores

Mean and standard deviation YRMS scores for each randomised group are reported for all visits (see Supplementary figure 5). The estimated difference between patients randomised to metyrapone and placebo at five weeks post randomisation (analysis of covariance of week 5 scores with baseline scores included as a covariate) was -0.04 (95% CI: -0.54, 0.46). Adjusting for origin of patient and variation between centres resulted in almost no change to the estimate; the adjusted estimate is -0.04 (95% CI: -0.52, 0.45). There is no evidence of different levels of manic symptoms between patients randomised to metyrapone and placebo.

Supplementary Figure 5: Young Mania Rating Scale (YMRS) scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.



Effect of metyrapone on state anxiety scores 5 weeks post randomisation

Based on an analysis of covariance the estimated effect of metyrapone relative to placebo at week 5 was an increase in state anxiety of 1.2 (95% CI: -0.7, 3.1). This is consistent with the larger fall in in anxiety between baseline and week 5 in the group randomised to metyrapone than in the other group that can be seen in supplementary figure 6. However the difference between groups is not statistically significant (p = 0.21). Adjusting for origin of patient care and differences between centres resulted in very little change in the estimated impact; adjusted estimate was an increase in anxiety of 1.2 (95% CI: -0.6, 3.0).

Supplementary Figure 6: State Trait Anxiety Inventory (STAI) - state scores of patients randomised to metyrapone or placebo over time. Data represents means with standard error bars.